Sensitivity Analysis of Data-driven Groundwater Forecasts to Hydroclimatic Controls in Irrigated Croplands

Alessandro Amaranto[1,2,3], Francesca Pianosi[4], Dimitri Solomatine[2,5,6], Gerald Corzo[2], Francisco Munoz-Arriola[1,7],

[1]Department of Biological Systems Engineering, University of Nebraska-Lincoln, Lincoln, Nebraska, United States.

[2]Hydroinformatics Chair Group, IHE Delft Institute for Water Education, Delft, The Netherlands.

[3]Department of Electronics, Information and Bioengineering, Politecnico di Milano, Milano, Italy.

[4]Department of Civil Engineering, University of Bristol, Bristol, UK.

[5]Water Resources Section, Delft University of Technology, Delft, The Netherlands

[6]Water Problem Institute of RAS, Moscow, Russia.

[7]School of Natural Resources, University of Nebraska-Lincoln, Lincoln, Nebraska, United States.

Corresponding author: Francisco Munoz-Arriola (fmunoz@unl.edu)

Key words: Groundwater forecasts, Artificial Neural Network, Uncertainty, Sensitivity Analysis

Abstract.

In the last decades, advancements in computational science have greatly expanded the use of artificial neural networks (ANNs) in hydrogeology, including applications on groundwater forecast, variable selection, extended lead-times, and regime-specific analysis. However, ANN-model performance often omits the sensitivity to observational uncertainties in hydroclimate forcings. The goal of this paper is to implement a data-driven modeling framework for assessing the sensitivity of ANN-based groundwater forecasts to the uncertainties in observational inputs across space, time, and hydrological regimes. The objectives are two-folded. The first objective is to couple an ANN model with the PAWN sensitivity analysis (SA). The second objective is to evaluate the scale- and process-dependent sensitivities of groundwater forecasts to hydroclimate inputs, computing the sensitivity index in groundwater wells (1) across the whole time-series (for the global sensitivity analysis); (2) across the output sub-regions with conditions of water deficit and water surplus (for the 'regional' sensitivity analysis); and (3) at each time step (for the time-varying sensitivity analysis). The implementation of the ANN-PAWN occurs in 68 wells across the Northern High Plains aquifer, USA, with pre-time-step rainfall, evapotranspiration, snowmelt, streamflow, and groundwater measurements as inputs. Results show that evapotranspiration and rainfall are the major sources of uncertainty, with the latter being particularly relevant in water surplus conditions and the former in water deficit conditions. The time-varying sensitivity analysis leads to the identification of localized sensitivities to other sources of uncertainty, as snowmelt in spring or river flow during the annual peak period at the groundwater level.

1 Introduction

In the past century, the growing access to pumping technologies and aquifer mapping has evidenced the role groundwater (GW) plays in securing food production and sustaining population growth (Konikow and Kendy, 2005). Agriculture consumes about 90% of the world's green water, and about 40% of irrigated water comes from groundwater withdrawals (Aeschbach-Hertig and Gleeson, 2012). The pressure exerted on global groundwater storage has led to global aquifers' depletion at rates of about 283 km3y-1 (Pokhrel et al., 2012), a value that represents an increase of 120% for the one observed in the 1960s (Wada et al., 2010).

Contrary to common perceptions, GW depletion is not limited to arid and semi-arid regions but also occurs in humid areas of the world. One of the best-documented cases is the High Plains aquifer (HPA) in the United States. The HPA, located in a temperate-subtropical area, has lost about 250 km3 of water in the past 60 years, corresponding to about 8% of the initial storage (Scanlon et al., 2012). Thus, effective water management is an unavoidable task, which could be achieved through a range of mechanisms, such as improved crop water use efficiency (Kukal and Irmak, 2017), irrigation scheduling (Kang et al., 2000) and reservoir operation optimization (Galelli et al., 2010).

In irrigated agriculture, water resources re-allocations are typically planned semi-seasonally or seasonally with the aim of optimizing water use efficiency, maintaining soil field capacity, and sustaining water systems (Amaranto et al., 2019). Hence, the successful implementation of seasonal water management strategies and irrigation scheduling relies on the ability to anticipate the future state of the GW system in response to various hydro-climatic and anthropogenic factors (Coppola et al., 2005). Data-driven models (DDMs) can be used for such forecasting purposes. DDMs are well-recognized techniques that extract the input-output relationship from data without requiring the complete characterization of a system. Developments of computational sciences have greatly expanded their application domain to hydrogeological systems, and DDMs have been used successfully for groundwater forecasts in many studies. One of the first applications of DDMs was implemented by Coulibaly et al., (2001), who tested and compared different ANN architectures for groundwater forecasting in Burkina Faso. A few years later, Daliakopulos et al. (2004) investigated the most suitable ANN architecture for predicting the GW level, finding that the most accurate model was a standard-feed forward neural network. More recent studies include Tapoglou et al. (2014), who simulated groundwater level variations across the Isar River using a combination of ANN and kriging (Bavaria, Germany). They found that this hybrid approach can be used successfully in aquifers, where the hydrogeological information is constrained. Mohanty et al. (2015) used

ANN to simultaneously forecast the weekly groundwater level at multiple sites, up to a maximum of a month. They found a significant decrease in performance for an increase in lead time. Barzegar et al. (2017) compared the ability of wavelet group data handling and extreme learning machines to forecast GW level three months ahead, concluding that the best performances can be obtained by the latter. Guzman et al. (2017) and Wunsch et al. (2018) forecasted daily GW level variations in a well in the Mississippi River Valley aquifer and Germany by using nonlinear autoregressive neural networks (NARX). Their results showed the potential of NARX to predict GW levels effectively. Amaranto et al. (2018) compared the ability of five different DDMs to forecast seasonal (1- to 4-month) GW levels across hydrological regimes. They found that the error of all the DDMs increased during intra-seasonal water-deficits. Amaranto et al. (2019) implemented an artificial neural network-instance based learning framework called Multi-Model Combination (MuMoC) to forecast GW levels in three hundred wells across the High Plains aquifer in response to irrigation demands and hydro-climatic inputs. The implementation of MuMoC led to finding that modeling performances were strongly affected by precipitation and evapotranspiration and that MuMoC outperformed and artificial neural network model in a single well, especially in areas where observations were abundant.

Nonetheless, DDMs do not require a complete hydrogeological characterization of the GW system, the performance of, for example, ANN models is sensitive to input measurements. Such discrepancies in the inputs can be attributed to operational errors, systematic bias, the geographical distance between weather stations and the monitoring wells, or the combination of the factors above. These observational uncertainties propagate through the model, leading to a decrease in predicting accuracy or a problematic interpretation of the results. The latter is more DDM-specific, given their intrinsic ‘black-box’ nature. In areas where GW is used for irrigation supply, and water allocation is scheduled ahead of time according to the projected water availability, it is critical to understand the dominant drivers of the GW model’s dynamics.

In other words, it is crucial to identify which variables need to be known with higher accuracy, and what effects the uncertainties of those variables have on the model outputs and forecast errors.

Thus, assessing the sensitivity of forecasting accuracy to observational uncertainty still represents a significant challenge for modelers and water managers, which can be addressed by global sensitivity analysis (GSA) techniques.

Modeling results might also be sensitive to different observational uncertainties (i.e., for different inputs) in different hydrogeological conditions (Corzo and Solomatine, 2007). A separate sensitivity analysis per each regime (hereafter referred to as 'regional' sensitivity analysis) is recommended. Usually, global sensitivity analysis methods use performance metrics aggregated over the whole simulation time series, which might lead to a significant loss of information regarding local behavior that might be of great interest (Pianosi and Wagner, 2015). Aggregating and performing SA at each time step (time-varying sensitivity analysis, TvSA) is a viable option for recovering significant sensitivity to input uncertainty at specific instants in time.

The goal for this study is to implement a framework for assessing a data-driven groundwater forecast (one month) sensitivity to multiple observational uncertainties in hydroclimate inputs (rainfall, evapotranspiration, snowmelt, river flow, and groundwater measurements) across space and time and for different hydrological regimes. The objectives are two folded. The first objective is to develop an ANN-based full-fledged framework, including an input-variable lag selection, and then we couple it with the global SA method called PAWN (Pianosi and Wagener, 2015). The second objective is to evaluate the scale- and process-dependent sensitivities of groundwater forecasts to hydroclimate inputs, computing the sensitivity index in groundwater wells (1) across the whole time-series (for the global sensitivity analysis); (2) across the output sub-regions with conditions of water deficit and water surplus (for the 'regional' sensitivity analysis); and (3) at each time step (for the time-varying sensitivity analysis).

The testbeds for the current experiment are 68 wells across the Northern High Plains aquifer. The authors carried deterministic analyses to characterize the spatial distribution of the error in groundwater forecasts in a previous study (Amaranto et al., 2019), which is not further discussed in this manuscript.

2 Methodology

2.1 Methodological Framework

To achieve the objectives described above, we apply the methodological framework outlined in Figure 1 to each of the wells selected for the analysis. In the first step, the hydroclimatic data (rainfall, evapotranspiration, river discharge, snowmelt, and groundwater level data) are divided into training and test sets (data division). Here, we optimize the split between the training set and the test set to ensure that both sets fit approximately the same statistical distribution, using the training and test average and standard deviation as optimization criteria. Then, the training minimum and maximum are used to normalize the data between 0 and 1 (data transformation). To select the most relevant lag times, we apply a model-based input variable selection (IVS) procedure (using Artificial Neural Networks as models) to the training set.

The training set is then further split into a proper training set and a cross-validation set. This procedure, referred here as cross-validation, is implemented to optimize the number of nodes in the ANN hidden layer, using the RMSE in the cross-validation set as criteria to be minimized. Unlike traditional applications of data-driven models, the test set is not just used to test the performance of the model but also to evaluate the sensitivity of the model's accuracy to input uncertainty. To implement this approach, we characterize each of the sources of uncertainty, and then we perform several perturbations on each of the inputs' time series (in the test set) accordingly. The perturbed input data are then iteratively sampled following a density-based sensitivity analysis scheme proposed by Pianosi and Wagener (2015), called PAWN. PAWN uses the difference between the conditional and the unconditional distributions of the output metric (RMSE in our case) to measure the sensitivity to different uncertain inputs. For each input combination sample, the ANN model is run, and the RMSE on the test set is used to evaluate the model's performance. Then, the difference between the unconditional and conditional distributions of the RMSE is used to compute the PAWN sensitivity indices in three conditions. First, to assess the overall effect of data uncertainties on model performance, the PAWN indices are computed for the RMSE calculated over the whole time series in each of the 68 wells under analysis. Second, to estimate the impact of data uncertainties in water deficit and surplus conditions, the PAWN indices are computed for the RMSE of the data-points below the 10% quantile of the water level hydrograph (deficit) and above the 90% quantile (surplus) (see Amaranto et al., 2018, for a more detailed description). Finally, to assess how the relative influence of different variables changes over time, we compute the PAWN indices for the RMSE evaluated at each time step with a moving window centered around the time step itself. Since the number of output time series in this paper is one per well (68 in total), for simplicity, the time-varying SA analysis is limited to two representative groundwater level time series. Further details about each of the blocks in Figure 1 are provided in the following sections.

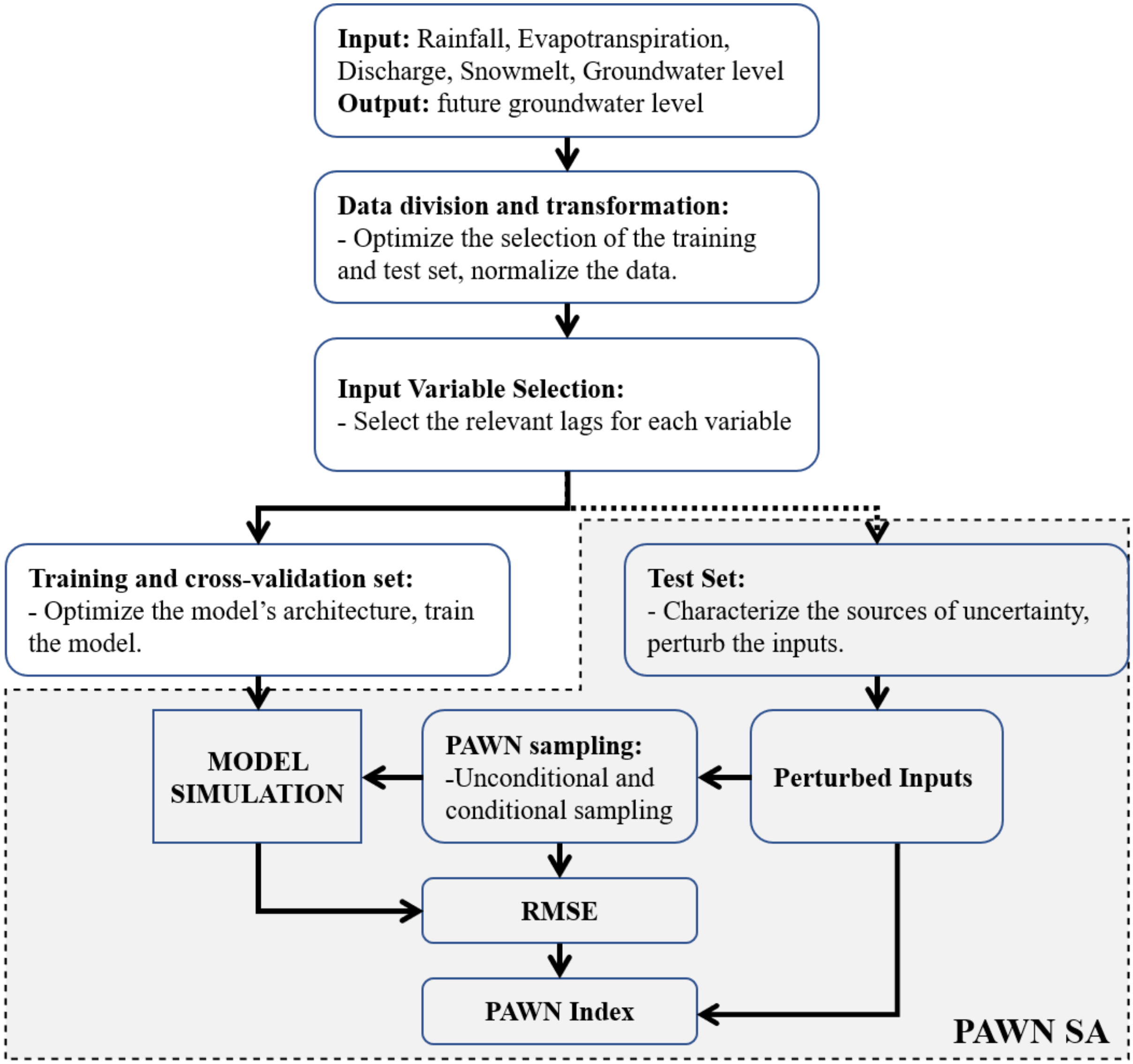

Figure 1: Methodological Framework employed in this study.

2.2 Data Division and Transformation

To assure that data come from the same population (Bhattacharya et al., 2007), the theory of DDM requires the statistical distributions of the training and the test sets to be approximately the same. First, we implement an iterative process of random selection to achieve the statistical homogeneity between the training and test sets. Then, we compare their distributions and select the split providing the closest statistical distribution. One drawback of this procedure is the inability to reproduce modeling results. In consequence, we chose to constrain the iterative randomization of the splits by limiting the search of the test set only to consecutive years, corresponding to 30% of the total number of time steps. For example, if we are supposed to have 30 years of data, the first nine years of the data are selected as the test set and the remaining 21 years of data as the training set. The statistical distributions of the two sets are compared using their mean and standard deviation, and the result is stored. In the second iteration, the test set is composed of the second-to-tenth year time-steps, and so on. The maximum statistical similarity is ensured by choosing the split(s) s* that satisfies the following rule:

$$s^* = \underset{s}{\operatorname{argmin}} \sqrt{(\mu_r(s) - 1)^2 + (\sigma_r(s) - 1)^2} \tag{1}$$

Where Where μ_R and σ_R are the ratios between means and standard deviations of the training and the testing set outputs (after normalization), respectively, and the optimal split(s) s* is selected by solving the Equation 1 through an exhaustive search procedure.

After selecting the optimum split, the minimum and the maximum of the training set are used to normalize the data in the interval [0-1]

2.3 Selection of Lags for the Input Variables

In building DDMs, a key step consists in the selection of relevant (and adequately lagged) input variables, a procedure commonly referred to as input variable selection (IVS). Often this is done by exhaustively testing all the possible combinations of properly lagged variables. However, due to the often-high number of candidates, the IVS procedure frequently becomes an optimization problem aimed at minimizing the trade-off between being computationally efficient (i.e., testing the least possible number of combinations) and finding the best input candidate (i.e., testing them all). Several studies have tried to address this problem. Among them, a genetic algorithm and general regression neural network (GAGRNN) proposed by Bowden (2005); a tree-based iterative search method developed by Galelli and Castelletti (2013); and a partial-mutual information-based algorithm (May et al. ,2008 and Elshorbagy et al., 2010a). A good variety of IVS methods is available in the literature (see, for example, Galelli et al., 2014, for a review). Considering our objective to evaluate the sensitivity of the groundwater forecasts to the uncertainties in the inputs, we include all the candidate input variables once in this study. Then, the problem is limited to selecting the proper lag for each input (rainfall, evapotranspiration, river discharge, and snowmelt).

To select the optimal lag for each variable, we perform a constrained ANN-based exhaustive search (CES). The CES algorithm iteratively tests any possible lag combination among the variables, each of them taken at one specific lag at the time. In other words, considering the four inputs mentioned above, and four lags (from *t* to *t-4*) per input, the CES generates 256 (4_4) potential input candidates. Each candidate includes rainfall, evapotranspiration, snowmelt, and streamflow (only referred to as flow from here on) once, in a lag going from *t* to *t-4*. For what concerns the fifth input (current groundwater level), we use only the last groundwater observation available (GW_t). This choice is based on the fact that, for this specific input, the lag 1 was the one maximizing the average mutual information with the model's output (GW_{t+1}). For each of the candidates, an ANN model is fitted on the training set. The RMSE in the cross-validation set was selected as optimization objective, to be minimized in the search of the best input subset.

2.4 Artificial Neural Networks

Multilayer perceptron (MLP, Haykin, 2004) neural networks are a machine learning technique that has been widely used in water-related studies (see, for example, Elshorbagy et al., 2010b; Abrahart et al., 2012). An MLP consists in an input layer, a hidden layer, and an output layer.

The first has the sole purpose of distributing the inputs further. The nodes in the hidden layer usually depend on the complexity of the system analyzed, but also on the number of input neurons. The number of nodes in the output layer is often one, or equal to the number of outputs. The connections between layers are associated with weights (w). A sigmoidal transfer function in the nodes of the hidden (and often of the output) layer(s) ensures the nonlinearity of the MLP.

2.5 Characterization of the Sources of Uncertainty

One of the objectives of this study is to assess the relative contribution of the uncertainties of the inputs on the accuracy of a data-driven model. Hence, the uncertainties in the observational inputs are divided into five categories: (1) the uncertainty in the rainfall observations, (2) the uncertainty in the evapotranspiration time series, (3) the uncertainty in snowmelt observations, (4) the uncertainty in streamflow time-series, and (5) the uncertainty in groundwater level observations used to both feed the model (autoregressive input) and evaluate it (output). Data uncertainty here is treated similarly, as in Pianosi and Wagener (2015). In particular, rainfall uncertainty was characterized, assuming that the measurement error is multiplicative, and the extent of the error changes differently in every rainfall event. This procedure, called storm-dependent rainfall depth multiplier, was first proposed and adopted by Kavetski et al. (2003, 2006). We assume a maximum observational rainfall error of ±40%. Therefore, the corresponding storm-dependent multipliers are extracted by a uniform distribution within the range [0.6, 1.4]. For evapotranspiration and snowmelt error, we assume a constant multiplier through the whole time series, drawing it from a uniform distribution over [0.7 - 1.3], i.e., assuming a maximum error of ±30%. These error percentages were decided by computing the average monthly coefficient of variation with respect to the climatology (defined here as the monthly cyclostationary average).

An additive error model was used to perturbing the flow data. Here, the errors are represented by a zero-mean autocorrelated heteroscedastic Gaussian process (HGp). The variance of the error model is considered linearly dependent on the flow (Schoups and Vrugt, 2010). The two parameters of this model are set to maintain the maximum error in flow observations at ±20% in 99% of the cases. Groundwater observations time series were treated similarly, but the HGp was fitted to the groundwater variations, rather than to the measurements themselves, to ensure that the measurement error is proportional to the difference in hydraulic head change, and not to its absolute value.

2.6 Evaluation Scheme

To evaluate the contribution of each input to the performance of the model, we use three different aggregation schemes of the forecasting errors. First, to identify the global contribution of the various inputs over time in each well in the study area, we compute the root mean squared error (RMSE) over the whole time series. Second, to assess the input importance in different hydrological conditions, we compute the RMSE over the region of the water levels above the upper (90%) and below the lower (10%) quantile of the water-level hydrograph. Finally, to assess the temporal evolution of the inputs relative influence, we compute the RMSE at each time step over a moving window centered on that time step:

$$RMSE_t = \sqrt{\frac{1}{2w + 1} \sum_{k = t-w}^{t+w} (gw_k^{sim} - gw_k^{obs})^2} \tag{2}$$

Where w is the semi-length of the moving window, t is the time step on which the window is centered, and gw_k^{sim} and gw_k^{obs} are respectively the simulated and observed groundwater levels on day k.

2.7 The PAWN Sensitivity Analysis

To assess the relative contribution of each input to the accuracy of the forecasts, we use a distribution-based sensitivity analysis method proposed by Pianosi and Wagener (2015) and called PAWN. The choice of this particular sensitivity method lies in its easy applicability to nonlinear models and its independence from the type of output distributions (for example, symmetric, multimodal, or highly skewed). Furthermore, it has shown to provide robust results for a relatively low sample size (Zadeh et al. 2017; Pianosi & Wagener, 2018). As other distribution-based methods, PAWN measures the sensitivity of the output y (the RMSE, in our case) to variations of an input x_i (the time-series of a particular hydrometeorological variable) by the distance between the unconditional distribution of y (obtained by varying all the inputs) and the conditional distribution obtained when all the inputs change but x_i. Here, the conditional and unconditional distributions are approximated by their empirical distribution functions. The distance between distributions is measured by the Kolmogorov-Smirnov statistic, computed as follows:

$$\mathbf{KS}(\mathbf{x_i}) = \max_{(\mathbf{y})} \left| \mathbf{F_y}(\mathbf{y}) - \mathbf{F_{y|x_i}}(\mathbf{y}|\mathbf{x_i}) \right| \tag{3}$$

where $F_y(y)$ is the empirical unconditional distribution of y, and $F_{y|x_i}(y|x_i)$ is the empirical conditional distribution of y when the *ith* input is kept fixed at the nominal value x_i. Since KS is dependent on such nominal value, the PAWN method considers KS statistics over a prescribed number of nominal values and then extracts their maximum as follows:

$$\mathrm{S_i} = \max_{(\mathrm{x_i})} \left[\mathrm{KS}(\mathrm{x_i})\right] \tag{4}$$

By definition, all the $KS(x_i)$ values, and consequently, the sensitivity indices S_i, vary in the range [0, 1]. The closer the unconditional distribution $F_y(y)$ is to the conditional ones $F_{y|x_i}(y|x_i)$, the smaller the $KS(x_i)$, values and therefore the smaller the sensitivity of y to x_i, and vice versa.

3 Experimental Setup

3.1 Artificial Neural Networks

To maximize the forecast performance, it is important to optimize the number of nodes in the hidden layer of the MLP. Here, the number of neurons was selected individually in each of the 68 wells under analysis within the interval [5, 17]. The MLP were trained by using the resilient backpropagation algorithm, using the R package RSNNS (Bergmeir and Benítez, 2012).

3.2 PAWN

As mentioned above, the PAWN index estimates the sensitivity of the model output to a given input by the difference between the unconditional and the conditional cumulative distribution functions (CDFs) of the output. The unconditional CDF is approximated here by the empirical distribution of N_u output samples obtained by sampling the whole input feasibility space. Similarly, the conditional CDFs are approximated by the empirical distributions of N_c output evaluations per each input. These evaluations require iterative sampling all the inputs but x_i, which is kept fixed to a nominal value. Since the index is dependent on the nominal value at which x_i is fixed, we repeat the evaluations using n different nominal values for x_i. Consequently, being M the number of variables, the total number of model evaluations required to compute the PAWN indices for M-inputs is $N_u + N_c \times n \times M$. The values of N_u, N_c, and n are fixed (by trial and error) to 5000, 3000, and 20, respectively, leading to a total number of model evaluations equal to 305,000 per well, and an average confidence interval size (obtained with 50 bootstraps) around the sensitivity index of 0.02.

The numerical implementation of the PAWN sampling and evaluation for our application is schematized in Figure 2. To obtain the unconditional distribution of y, we randomly sample each of the input factors 5000 times. Each of these 5000 samples corresponds to a dataset containing one perturbed time series of rainfall, evapotranspiration, snowmelt, discharge, and current GW level. These input datasets are fed iteratively into the ANN model, which will, therefore, produce 5000 time series of GW level forecasts. Then, by comparing GW forecasts and observations, we obtain 5000 realizations of the model performances (i.e., 5000 values of RMSE, or 5000 RMSE values at each time step in case of TvSA), which are used to approximate the unconditional distribution.

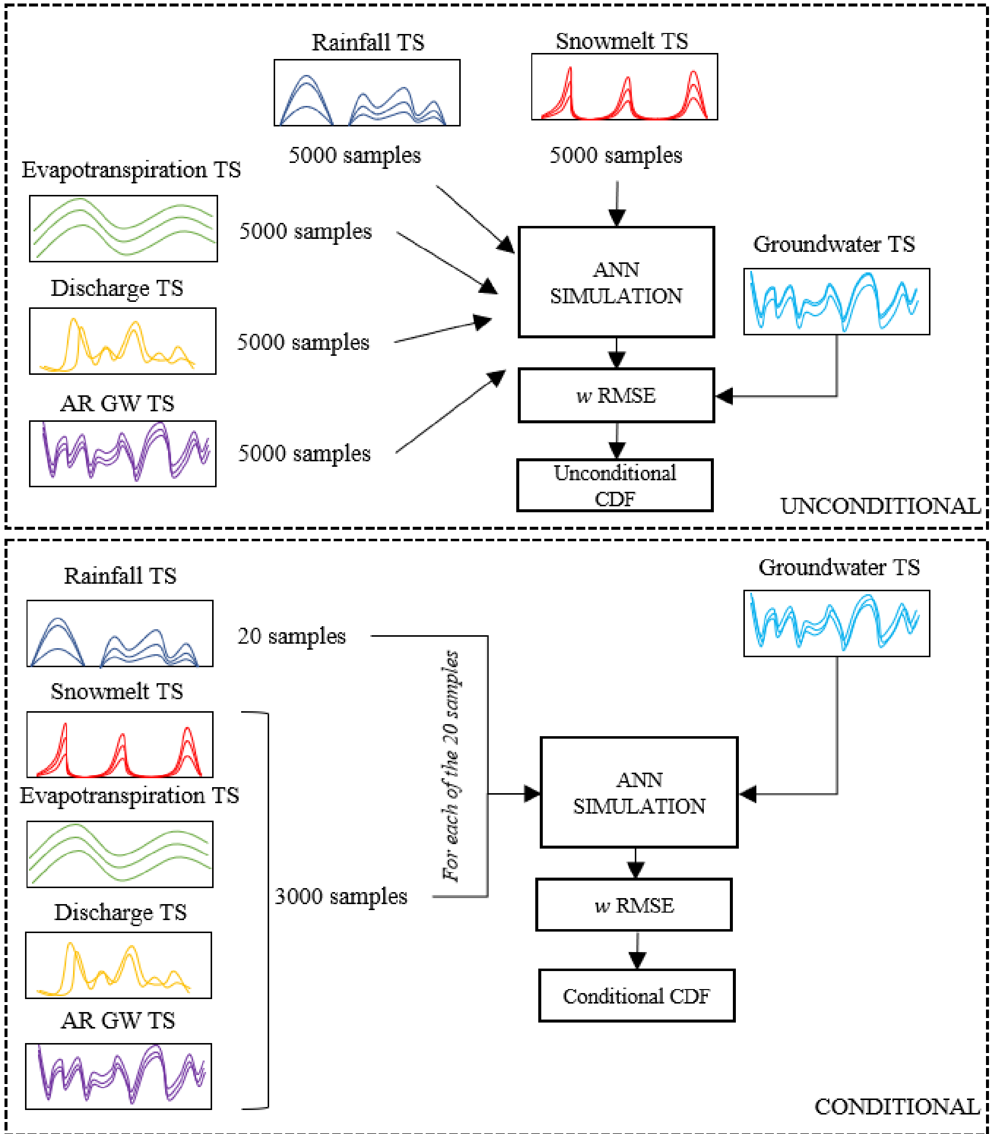

Figure 2: PAWN experimental setup (TS stands for time-series; ARGW TS is the Autoregressive term of groundwater level time-series).

The steps required for the numerical approximation of the conditional distributions are represented in the bottom part of Figure 2. For the sake of simplicity, Figure 2 refers to only one of the inputs (in this case, rainfall), but the procedure for the other inputs remains the same. First, we randomly sample one conditional rainfall time series. Then, we generate 3000 random samples of the other time series, and we iteratively run the model (in this case, the rainfall time series is fixed while snowmelt, discharge, evapotranspiration, and GW level time series change at each of the 3000 iterations). The 3000 RMSE values associated with the model forecasts time series are then used to approximate the conditional distributions. Then, we apply Equation 3 to compute the KS statistic, we rerun the experiment as many times as the number of conditioning values (20 in the current analysis), and we compute the PAWN index as in Equation 4. To achieve the specific objectives of this study, we compute the PAWN indices for the RMSE calculated over (1) the whole time series; (2) water scarcity and abundance conditions; and (3) at each time step using a window semi-length of three months (w = 3 months). Also, a six-month window is tested.

The PAWN analysis is implemented using an R adaptation of the SAFE Toolbox (Pianosi et al., 2015).

4 Material

4.1 Case Study and Dataset

The study area in the High Plains aquifer (HP, Figure 3a) extends for about 450,000 km^2 (the largest aquifer in the United States) over eight states (South Dakota, Nebraska, Colorado, Kansas, Oklahoma, Wyoming, New Mexico, and Texas). Since the 1950s, the aquifer has been intensively exploited by irrigation, and now ranks first in the United States for groundwater withdrawal. In the last 30 years, water levels in the HP have shown declines of more than 30 m. These declines caused a saturated thickness reduction in some areas (Kansas and Texas, in particular) of more than 50% (Scanlon et al., 2012). The total GW depletion in the HP in the past 70 years is about 8% of the total groundwater storage.

The area under investigation is the Northern portion of the High Plains (Figure 3b-e), which occupies about 37% (167,000 km^2) of the total aquifer area. It is crossed by the Platte River, which drains northeast Colorado, southeast Wyoming, and central Nebraska before merging into the Missouri River (Eschner et al., 1983). Here, the aquifer is constituted by unconsolidated Quaternary alluvial deposits and is mainly in unsaturated conditions, with total saturated thickness ranging from 400 m in the central part to less than 50 m in the west (McGuire, 2017).

Irrigation (measured in terms of percentage of irrigated area, Ozdogan and Gutman, 2008) is particularly developed in the eastern part and alongside the Platte River (Figure 3b), with corn and soybeans being the most cultivated crops. The irrigation system is usually a center pivot sprinkler. According to Wen and Chen (2006), the number of registered irrigation wells grew from 1200 in 1936 to about 100,000 in 2007, serving about 85% of the state's irrigation land.

Rainfall (Figure 3d) follows a west-to-east gradient with a minimum of about 27 mm/month near the border with Wyoming to a maximum of about 70 mm/month on the eastern side of the aquifer. The maximum net recharge-rate of the aquifer occurs in the east part of Nebraska (mainly rainfall-driven) and alongside the Platte River, and it is of about 22 mm/y (Houston et al., 2013). The contribution of the Platte River to aquifer recharge is also evident from Figure 3e, where it is possible to observe how the area close to the river is the one characterized by the highest rise in the GW level in the past 70 years. GW level increases are also frequent in the north-central part of the state, where low irrigation intensity and high saturated aquifer thickness might be considered the main drivers of the aquifer recharge. Water level decrease is particularly severe in the southeast and in the southwest.

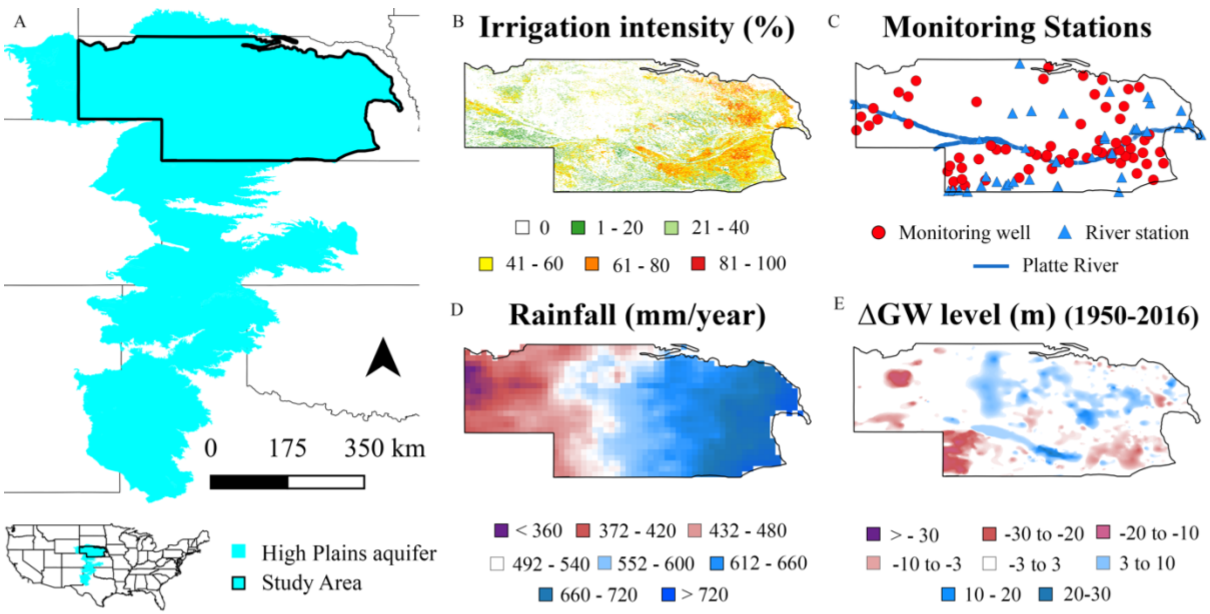

Figure 3: (A): location of the High Plains aquifer and of the study area; (B) irrigation intensity (percentage of irrigated areas, Ozdogan and Gutman, 2008); (C) location of the wells under analysis and of river discharge monitoring stations; (D) Annual rainfall (Rodell et al., 2004); (E) Decrease in water table level in the period 1950-2016.

Monthly estimation of rainfall (P, mm/month), evapotranspiration (mm/month), and snowmelt (mm/month) were obtained by the Global Land Data Assimilation System (GLDAS, Rodell et al., 2004) with a spatial resolution of 1/8-degree latitude x longitude (about 15 x 15 km). GW (in meters below land surface) and discharge (Q, m^3/d) in the HP aquifer data were provided by the USGS (2015). We filtered the complete USGS GW database to exclude stations with an observation period of fewer than ten years of data (120 observations) and missing data higher than 25% within the 1980-2018 period. After the implementation of the filter, 68 wells remained available for analysis (Figure 3c). Streamflow data were gathered from the stream gauges closest to the selected monitoring wells.

5 Results and Discussion

5.1 Spatial Global Sensitivity to Data Uncertainty

Figure 4 shows the spatial distribution of the sensitivity index for each of the five variables assessed in this study. By looking at the chart and in Table 1, it is easy to notice the strong impact that rainfall and evapotranspiration uncertainties have on ANN performances. In contrast, the contribution of snowmelt is practically negligible. One possible explanation for this might lie in the fact that, while Figure 4 shows aggregated results for the whole time series, snowmelt is a phenomenon that usually occurs only a few months a year (in February, March, and April, see Amaranto et al., 2019 for additional elements). Its contribution is limited to this time frame. Therefore, while its impact on the model's performances in a time step might be relevant, its overall contribution appears to be much lower. Also, the interaction of snowmelt with the upper soil layers is well known, and it is unlikely that, in locations where the aquifer is deeper, this variable might have any influence on groundwater dynamics.

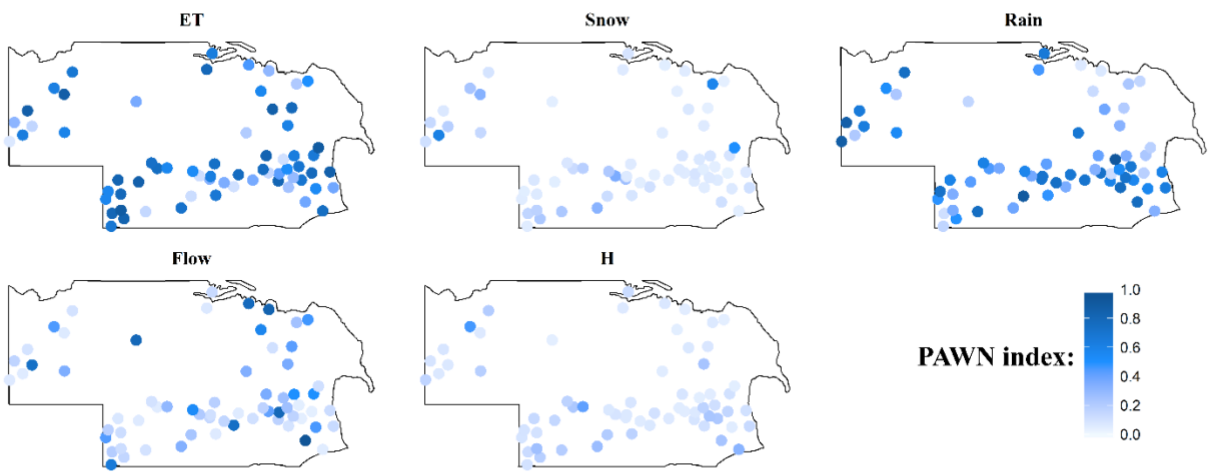

Figure 4: Spatial distribution of the PAWN sensitivity index computed for each input variable ET = evapotranspiration; Snow = snowmelt, Rain = rainfall; Flow = streamflow and H = groundwater level measurement at previous time-step.

Table 1: Mean, maximum, and minimum value of the PAWN index across the study area

	ET	Snow	Rain	Flow	H
mean(PAWN)	0.56	0.12	0.49	0.30	0.08
max(PAWN)	0.89	0.62	0.89	0.90	0.42
min(PAWN)	0.07	0.03	0.06	0.04	0.04

Figure 5 shows the variables producing the highest and the second-highest value of the PAWN index in each of the wells analyzed. Analyzing Figure 4, Figure 5, and Table 1, one can see that, overall, evapotranspiration (μPAWN = 0.56), rainfall (μPAWN = 0.49), and river flow (μPAWN = 0.3) are the three dominant variables governing model performances. In particular, evapotranspiration was the most relevant variable in 37 wells (54% of the cases) and the second most relevant in another 19 wells (27% of the total), followed by rainfall (the most pertinent input in 21 wells, 31% of the whole; and the second most relevant in 30, 40% of the total) and river flow (most relevant input in 9 wells, 13% of the total, second-most appropriate in 12 wells, 17% of the total).

By comparing Figure 5 and Figure 3b, we see that evapotranspiration uncertainties seem to mainly affect the performance of the models in regions where irrigation intensity is higher (orange and red areas in Figure 3b). The influence of flows can be more robust near rivers, but flow measurement stations were not always available near wells to effectively couple the discharge time series with the groundwater levels. On the other hand, the influence of rainfall on groundwater level changes can be particularly relevant along the Platte River.

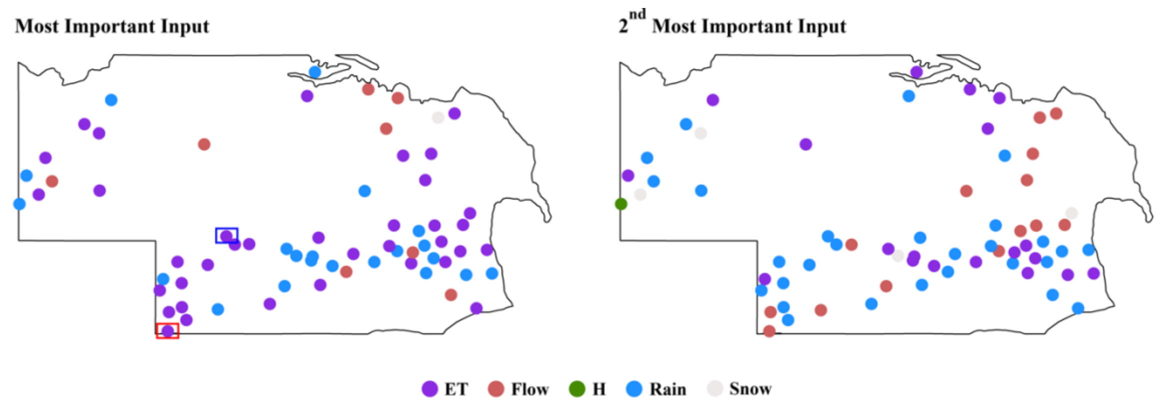

Figure 5: Variable producing the highest (left panel) and second highest (right panel) sensitivity index in each of the 68 wells. The blue and red rectangles represent the wells selected for time varying SA.

5.2 Regional sensitivity analysis for water availability regimes

Figure 6 shows the input variables responsible for the highest uncertainty in forecasts during water surplus (left panel) and water deficit (right panel) conditions. By looking at the figure on the left, one can notice the increased relevance of snowmelt, rainfall, and flow. This close relationship between surface water-based variables and groundwater levels is probably because the upper quantile corresponds to the hydrograph section associated with the water level peak, usually occurring between February and April. During those months, snowmelt occurs and recharges the aquifer. As a consequence, snowmelt becomes the most relevant input variable in water abundance conditions in six of the wells under analysis, a situation in which the overall sensitivity analysis never occurred. March and April are also the months when maximum rainfall usually occurs and when forecast sensitivity to precipitation uncertainty is the most relevant in 26 of the wells. Besides, the higher water level in the upper quantile favors river seepage (which is inversely proportional to the distance between river sediment and groundwater level), and consequently, sensitivity to flow data uncertainty increases, with flow being the most important source of uncertainty in 16 wells. As expected, the left panel in Figure 6 also shows how relevant is the decrease in evapotranspiration when there is a water surplus. For example, the relationship between evapotranspiration and crop water demand, and it is maximum during the crop-growing season, is more evident later on in the year, causing a significant intra-annual water-level depletion.

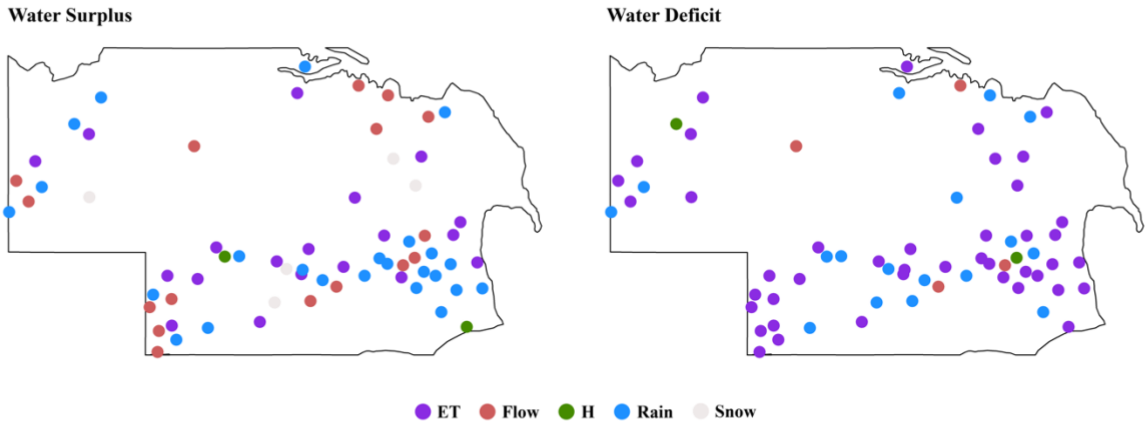

Figure 6: Most important input factor in water surplus (left panel) and in water deficit (right panel) conditions.

At the same time, Figure 6 shows how evapotranspiration is by far the primary source of forecast uncertainty in the lower quantile of the water level hydrograph. Overall, 44 out of the 68 wells (about 65% of the total wells assessed) had ET associated with the highest PAWN value. As stated above, ET is at its maximum during the crop growing season, when significant GW depletions also occur. In particular (and as we will see in the following sections), the peak in ET usually occurs in August, which is also the month corresponding to the yearly minimum in groundwater level and the maximum drawdown. Consequently, uncertainty in evapotranspiration inputs can propagate from ET to the forecasts of groundwater levels. This propoagation is more evident during months in the lower quantile, and when the forecast sensitivity to evapotranspiration becomes the most relevant among all inputs analyzed in this study.

5.3 Time-varying Sensitivity to Input Data Uncertainty

Regarding the temporal variability of the PAWN index, Figure 7 shows the time series (February 1991-October 2001) of the GW level (red line in the plot) and the PAWN index (grayscale rectangles) in one of the monitoring wells (MW1, red box in Figure 5a). The location of MW1 is near the Lower Republican River, in the southern part of Nebraska (Figure 3c). In MW1, the aquifer is relatively shallow (the average groundwater depth is 2 m), allowing surface water and groundwater to interact. The initial portion of the time series shows a keen sensitivity of flow observational uncertainties on modeling error, with flow influence being particularly relevant during the rising limb of the water table level hydrograph. As can be seen in the figure, snowmelt has a periodical control, with peaks on the PAWN index regularly occurring between February and April, when snowmelt occurs. This influence seems to confirm the previous finding that, despite the low overall sensitivity to snowmelt, there are instances in time when this variable at least marginally influences modeling performances. However, snowmelt influence dissipates in the second half of the time series (from 1997), when the pattern in groundwater levels also changes. Staring in 1997, groundwater depletion during the growing season appears to be much more acute (on average, five times greater than the depletion rates occurring between 1991 and 1996). This increased depletion might cause groundwater level changes occurring deeper from the surface in the spring, reducing the effect of snowmelt on the model error. At the same time, the model exhibits an increase in the sensitivity to evapotranspiration during the crop-growing season. The best possible explanation for this period is an increase in groundwater use for irrigation. In essence, crop irrigation requirements (and consequently evapotranspiration) govern the groundwater variability in the season when irrigation takes place. Hence, an increase in irrigation water use might lead to more considerable influence of evapotranspiration uncertainties on modeling performances.

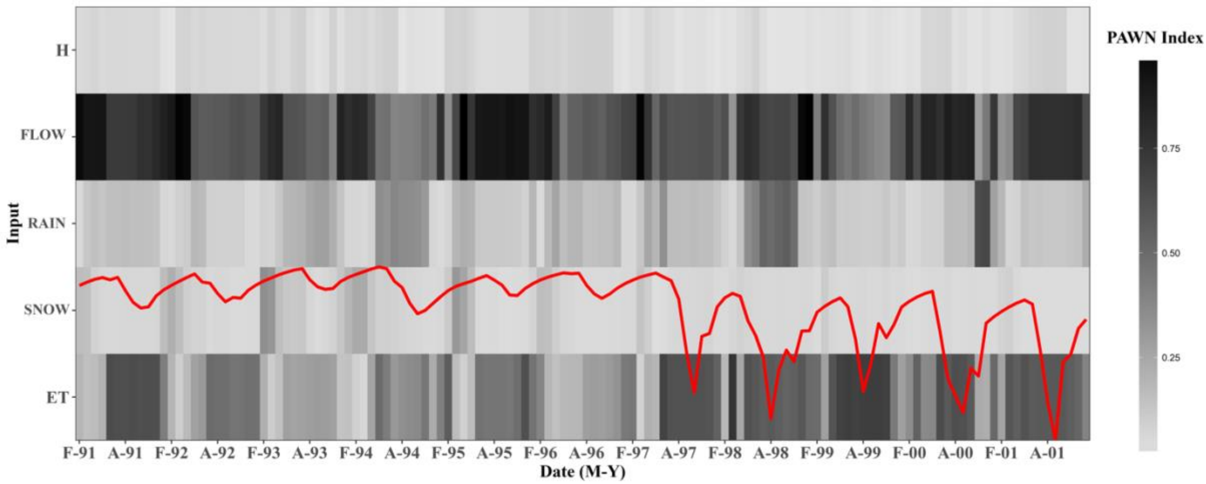

Figure 7: Time-varying PAWN index in Monitoring Well 1 (the red line is a qualitative representation of the normalized GW level changes).

Figure 8 shows the time-varying PAWN index for MW2 (in the blue box of Figure 5a). As in MW1, snowmelt likely influences the strong seasonality in the figure. However, unlike in the previous case, the influence of river flow on MW2 appears to be more seasonal rather than a continuous effect along with the time series. The deeper water level might explain this effect in MW2, which varies from a minimum of about 10 m in March and April to a maximum of about 19 m in August and September, in comparison to MW1's shallow groundwater level. The only time when any interaction between the surface water and groundwater emerges is when the spring recharge might be responsible for bringing the water table level closer to the surface. Practically no interaction between the two occurs through the rest of the year. In the case of MW2, the performance of the model looks to be entirely driven by rainfall and evapotranspiration, with the latter showing an increasing influence in the second portion of the time series (between March 2000 and December 2005). As in the previous case, the increased influence of evapotranspiration coincides with much deeper water tables during the growing season. For instance, after the summer of 2000, the water level experienced a drastic depletion in the water table during summer, which decreased the autumn-spring recovery typically observed in the previous five years. Furthermore, the water level starts showing a low depletion trend during which the influence of rainfall decreases, and the influence of evapotranspiration consistently increases.

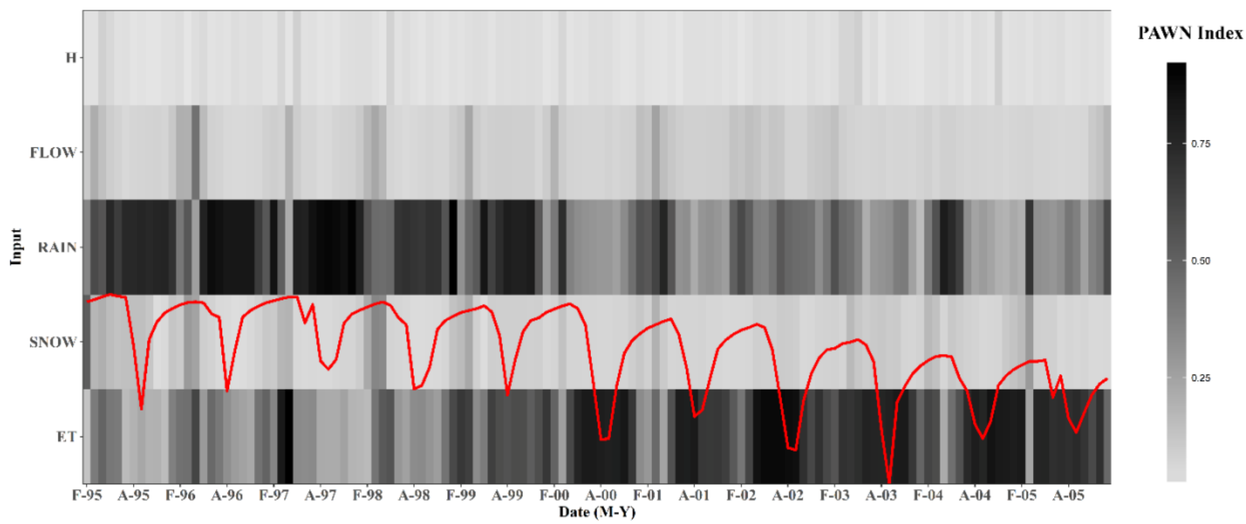

Figure 8: Time-varying PAWN index in Monitoring Well 2 (the red line is a qualitative representation of the normalized GW level changes).

5.3.1 Effect of changing the window size

Figure 9 illustrates an unclear increase in the window in MW2 for the sensitivity of groundwater changes to rainfall and evapotranspiration when w = 6 months. The time series has two sections, one section (1995-2000), predominantly rainfall-driven, and another section (2000-2005) evapotranspiration-driven.

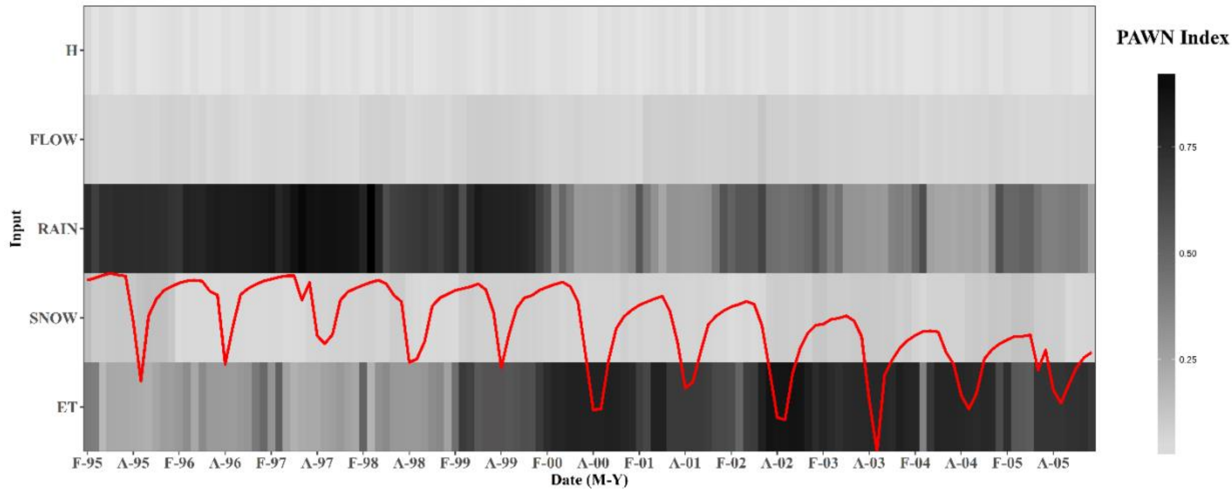

Figure 9: Time-varying PAWN in Monitoring Well 2 for window semi-length w = 6 months (the red line is a qualitative representation of the normalized GW level changes).

Also, Figure 9 indicates that the effect of snowmelt and flow becomes practically negligible throughout the time series. This result might be explained by the fact that both variables have a significant impact on modeling results only for limited and specific times. The effect of flow was relevant only around March-April, while the snowmelt effect was detectable only around February-March. These months also correspond to the only time of the year when snowmelt (2 mm/day) is comparable to rainfall (1.8 mm/day). By increasing the window length, the estimated sensitivity index for those months contrasts with the low sensitivity obtained in the months before February and after April. Thus, an apparent combination of conditions makes the contribution of flow and snowmelt practically undetectable. At the same time, rainfall and evapotranspiration lead to a more regular sensitivity index (characterized by fewer variations between one-time step and the following). In essence, drastic changes in the PAWN index, such as the one occurring for evapotranspiration in March and April 2007 (or the one for rainfall in March 2000), are attenuated and become practically negligible.

6 Conclusions

In this study, we implemented a SA framework to better understand the sensitivity of ANN errors to input observational uncertainties in groundwater forecast.

As a product of the coupling ANN-SA, we conclude the following:

Overall, evapotranspiration (μPAWN = 0.56) and rainfall (μPAWN = 0.49) were the most relevant inputs. In particular, evapotranspiration appeared to be particularly relevant in areas with higher irrigation intensity, whereas the rainfall effect was detectable, especially in the Platte River area. Modeling errors were not sensitive to the groundwater level measurement error in any of the case studies.

Results for flow were difficult to interpret since flow stations were unavailable for coupling with the time series at all 68 wells. However, the flow effect was higher in the geographic proximity to the Platte and Lower Republican rivers.

The contribution of snowmelt to the changes in groundwater levels was practically negligible across the studied area (average PAWN index = 0.12). Two factors might drive this effect. The first factor is that snowmelt occurs one to two months in any given year, and the second factor is that the performance of the model might be relevant in a single time step, but the effect is much lower throughout the whole time series.

Regional SA results showed that evapotranspiration is the most relevant variable in water scarcity conditions (10% quantile of the water level hydrograph). It showed in fact to dominate the error dynamics in about 65% of the wells in the study area. In contrast, rainfall was the most important in water surplus (90% quantile); being the major sources of uncertainty in 40% of the analyzed wells. Sensitivity to snowmelt and flow also showed an increase in the upper quantile.

The time-varying SA was able to register information that otherwise would have been lost by applying SA to the whole time series. For example, the analysis of the constrained window shows that the effect of snowmelt is significant at the beginning of spring, with peaks of sensitivity index up to 0.62. Also, evapotranspiration proved influential in seasons when the groundwater depletion was particularly severe, while at other times, flow or rainfall was the most relevant variables.

Increasing the window size led to less variability in the results and, consequently, to a less qualitative interpretation. Additionally, it hides potentially relevant information, such as the effect of snowmelt and river seepage in the spring.

In summary, the present study evidence how complex phenomena govern the ANN ability to predict GW availability in irrigated areas in the land surface and the subsurface and across different spatial, hydrological, and temporal scales. Accurate estimations of evapotranspiration are critical since it was identified as the primary source of uncertainty in the forecast of groundwater levels. Furthermore, regional and time-varying sensitivity analyses --tailored for specific water regimes-- were able to identify the importance of other forcing inputs (e.g., rainfall in water surplus, and snowmelt at the beginning of the year), which could not be captured when those errors were averaged over the entire time-series. These analyses are recommended in order to raise awareness of the multiple sources of uncertainty and their roles in governing specific hydrological conditions and during particular periods.

7 Limitations and Future Recommendations.

This study is limited by the lack of real-world pumping data (which were not available for the case study area) and by the use of proxies, such as evapotranspiration, to simulate crop water requirements. Using pumping data would have provided more information on how human intervention shapes model performance. Furthermore, the selection of the feasibility space for the perturbed input was empirically established. When possible, this choice should be made based on information about the error (available, perhaps, from local institutions). The analyses of the TvSA indicate how a different window convey different information. A suggestion is to investigate various sizes, to capture the full range of sensitivities across time-scales. Also, SA results might be sensitive to the choice of the model. Here, we used artificial neural networks to forecast GW levels and GSA to estimate the effect of data uncertainty on the model's performance. The choice of a different model (perhaps physically-based) might lead to different results. The use of a physically-based model (coupled with an analysis not based on error metrics such as the presented here) might likewise provide insights on how the physical system (and not the model's error) is sensitive to uncertainties in forcings and parameters. Therefore, further research on coupling physically- and data-driven models should might lead estimate the contributions of the multiple sources of uncertainty in sub-seasonal forecasts of groundwater levels.

8 Acknowledgments

The authors acknowledge the support provided by the Robert B. Daugherty Water for Food Global Institute at the University of Nebraska and the University of Nebraska-Lincoln Institute of Agriculture and Natural Resources-Agricultural Research Division. Some research ideas and components were also developed within the framework of the USDA National Institute of Food and Agriculture, Hatch project NEB-21-166 Accession No.1009760 and NEB-21-176 Accession No. 1015252, and grant no. 17-77-30006 of the Russian Science Foundation. Post-processed input data and experiments results can be founded in (https://github.com/alessandroamaranto/HP_SA).

9 References

Abrahart, R. J., Anctil, F., Coulibaly, P., Dawson, C. W., Mount, N. J., See, L. M., Shamseldin, A. Y., Solomatine, D.P., Toth, E., and Wilby, R. L.: Two decades of anarchy? Emerging themes and outstanding challenges for neural network river forecasting. Progress in Physical Geography: Earth and Environment, 36(4), 480-513, https://doi.org/10.1177/0309133312444943, 2012.

Aeschbach-Hertig, W., and Gleeson, T.: Regional strategies for the accelerating global problem of groundwater depletion. Nature Geoscience, 5(12), 853-861, 2012.

Amaranto, A., Munoz-Arriola, F., Corzo, G., Solomatine, D. P., and Meyer, G. E.: Semi-seasonal groundwater forecast using multiple data-driven models in an irrigated cropland. Journal of Hydroinformatics. 20(6), 1227-1246, https://doi.org/10.2166/hydro.2018.002, 2018.

Amaranto, A., Munoz-Arriola, F., Solomatine, D. P., and Corzo, G.: A spatially enhanced data-driven multi-model to improve semi-seasonal groundwater forecasts in the High Plains aquifer, USA. Water Resources Research, 55(7), 5941-5961. https://doi.org/10.1029/2018WR02430, 2019.

Barzegar, R., Fijani, E., Moghaddam, A. A., and Tziritis, E: Forecasting of groundwater level fluctuations using ensemble hybrid multi-wavelet neural network-based models. Science of the Total Environment, 599-600, 20-31, https://doi.org/10.1016/j.scitotenv.2017.04.189, 2017.

Bergmeir, C., and Benítez, J. M: Neural networks in R using the Stuttgart Neural Network Simulator: RSNNS. Journal of Statistical Software, 46(7), 2012.

Bhattacharya, B., Price, R. K., and Solomatine, D. P: A machine learning approach to modeling sediment transport. Journal of Hydraulic Engineering, 133(4), 440-450, https://doi.org/10.1061/(ASCE)0733-9429(2007)133:4(440), 2007.

Bowden, G. J., Maier, H. R., and Dandy, G. C: Input determination for neural network models in water resources applications. Part 2. Case study: forecasting salinity in a river. Journal of Hydrology, 301(1-4), 93-107, https://doi.org/10.1016/j.jhydrol.2004.06.020, 2005.

Coppola, E. A., Rana, A. J., Poulton, M. M., Szidarovszky, F., and Uhl, V. W.: A neural network model for predicting aquifer water level elevations. Groundwater, 43(2), 231-241, https://doi.org/10.1111/j.1745-6584.2005.0003.x, 2005.

Corzo, G., and Solomatine, D.: Baseflow separation techniques for modular artificial neural network modelling in flow forecasting. Hydrological Sciences Journal, 52(3), 491-507, https://doi.org/10.1623/hysj.52.3.49, 2007.

Elshorbagy, A., Corzo, G., Srinivasulu, S., and Solomatine, D. P.: Experimental investigation of the predictive capabilities of data driven modeling techniques in hydrology - Part 1: Concepts and methodology. Hydrology and Earth System Sciences, 14(10), 1931-1941, DOI: 10.5194/hess-14-1931-2010, 2010a.

Elshorbagy, A., Corzo, G., Srinivasulu, S., and Solomatine, D. P.: Experimental investigation of the predictive capabilities of data driven modeling techniques in hydrology - Part 2: Application. Hydrology and Earth System Sciences, 14(10), 1943-1961, doi:10.5194/hess-14-1943-2010, 2010b.

Eschner, T.: Hydrologic and Geomorphic Studies of the Platte River Basin (Professional Paper 1277). Washington, DC: U.S. Government Printing Office, 1983.

Galelli, S., and Castelletti, A.: Tree-based iterative input variable selection for hydrological modeling. Water Resources Research, 49(7), 4295-4310, https://doi.org/10.1002/wrcr.20339, 2013.

Galelli, S., and Soncini-Sessa, R.: Combining metamodelling and stochastic dynamic programming for the design of reservoir release policies. Environmental Modelling & Software, 25(2), 209-222, https://doi.org/10.1016/j.envsoft.2009.08.001, 2010.

Galelli, S., Humphrey, G. B., Maier, H. R., Castelletti, A., Dandy, G. C., and Gibbs, M. S.: An evaluation framework for input variable selection algorithms for environmental data-driven models. Environmental Modelling & Software, 62, 33-51, https://doi.org/10.1016/j.envsoft.2014.08.015, 2014.

Guzman, S. M., Paz, J. O., and Tagert, M. L. M.: The use of NARX neural networks to forecast daily groundwater levels. Water Resources Management, 31(5), 1591-1603, doi: 10.1007/s00500-015-1833-z, 2017.

Haykin, S.: Neural Networks: A comprehensive foundation. Neural Networks, 41, 2004.

Houston, N. A., Gonzales-Bradford, S. L., Flynn, A. T., Qi, S. L., Peterson, S. M., Stanton, J. S., Sohl, T.L., and Senay, G. B.: Geodatabase compilation of hydrogeologic, remote sensing, and water-budget-component data for the High Plains aquifer, 2011 (Data Series 777). Reston, VA: U.S. Geological Survey. Retrieved from https://pubs.usgs.gov/ds/777/pdf/ds777.pdf, 2013.

Kang, S., Shi, W., and Zhang, J.: An improved water-use efficiency for maize grown under regulated deficit irrigation. Field crops research, 67(3), 207-214, https://doi.org/10.1016/S0378-4290(00)00095-2, 2000.

Kavetski, D., Franks, S. W., and Kuczera, G.: Confronting input uncertainty in environmental modelling. Calibration of Watershed Models, 6, 49-68, https://doi.org/10.1029/WS006p0049, 2003.

Kavetski, D., Kuczera, G., and Franks, S. W.: Bayesian analysis of input uncertainty in hydrological modeling: 1. Theory. Water Resources Research, 42(3), DOI: 10.1029/2005WR004368, 2006.

Konikow, L. F., and Kendy, E.: Groundwater depletion: A global problem. Hydrogeology Journal, 13(1), 317-320, DOI: 10.1007/s10040-004-0411-8, 2005.

Kukal, M. S., and Irmak, S.: Spatial and temporal changes in maize and soybean grain yield, precipitation use efficiency, and crop water productivity in the us great plains. Transactions of the ASABE, 60(4), 1189-1208, DOI: 10.13031/trans.12072, 2017.

May, R. J., Maier, H. R., Dandy, G. C., and Fernando, T. G.: Non-linear variable selection for artificial neural networks using partial mutual information. Environmental Modelling & Software, 23(10-11), 1312-1326, https://doi.org/10.1016/j.envsoft.2008.03.007, 2008.

McGuire, V. L.: Water-level and recoverable water in storage changes, High Plains aquifer, predevelopment to 2015 and 2013–15 (Scientific Investigations Report 2017-5040). Reston, VA: U.S. Geological Survey. Retrieved from http://pubs.er.usgs.gov/publication/sir20175040, 2017.

Mohanty, S., Jha, M. K., Raul, S. K., Panda, R. K., and Sudheer, K. P.: Using artificial neural network approach for simultaneous forecasting of weekly groundwater levels at multiple sites. Water Resources Management, 29(15), 5521-5532, https://doi.org/10.1007/s11269-015-1132-6, 2015.

Ozdogan, M., and Gutman, G.: A new methodology to map irrigated areas using multi-temporal MODIS and ancillary data: An application example in the continental US. Remote Sensing of Environment, 112(9), 3520-3537, https://doi.org/10.1016/j.rse.2008.04.010, 2008.

Pianosi, F., Sarrazin, F., and Wagener, T.: A Matlab toolbox for global sensitivity analysis. Environmental Modelling & Software, 70, 80-85, https://doi.org/10.1016/j.envsoft.2015.04.009, 2015.

Pianosi, F., and Wagener, T.: A simple and efficient method for global sensitivity analysis based on cumulative distribution functions. Environmental Modelling & Software, 67, 1-11, https://doi.org/10.1016/j.envsoft.2015.01.004, 2015.

Pianosi, F., & Wagener, T.: Understanding the time-varying importance of different uncertainty sources in hydrological modelling using global sensitivity analysis. Hydrological Processes, 30(22), 3991-4003. 2016

Pokhrel, Y. N., Hanasaki, N., Yeh, P. J., Yamada, T. J., Kanae, S., and Oki, T.: Model estimates of sea-level change due to anthropogenic impacts on terrestrial water storage. Nature Geoscience, 5(6), 389-392, https://doi.org/10.1038/ngeo1476, 2012.

Rodell, M., Houser, P. R., Jambor, U., Gottschalck, J., Mitchell, K., Meng, C.-J., Arsenault, K., Cosgrove, B., Radakovich, J., Bosilovich, J., Walker, J.P., Lohmann, D., and Toll, D.: The Global Land Data Assimilation System. Bulletin of the American Meteorological Society, 85(3), 381-394, https://doi.org/10.1175/BAMS-85-3-381, 2004.

Scanlon, B. R., Faunt, C. C., Longuevergne, L., Reedy, R. C., Alley, W. M., McGuire, V. L., and McMahon, P. B.: Groundwater depletion and sustainability of irrigation in the US High Plains and Central Valley. Proceedings of the National Academy of Sciences, 109(24), 9320-9325, https://doi.org/10.1073/pnas.1200311109, 2012.

Schoups, G., and Vrugt, J. A.: A formal likelihood function for parameter and predictive inference of hydrologic models with correlated, heteroscedastic, and non-Gaussian errors. Water Resources Research, 46(10), https://doi.org/10.1029/2009WR008933, 2010.

Tapoglou, E., Karatzas, G. P., Trichakis, I. C., and Varouchakis, E. A.: A spatio-temporal hybrid neural network-Kriging model for groundwater level simulation. Journal of Hydrology, 519, 3193-3203, https://doi.org/10.1016/j.jhydrol.2014.10.040, 2014.

USGS: National Water Information System: USGS Groundwater Data for the Nation. https://waterdata.usgs.gov/nwis/gw (accessed 14 August 2018), 2015.

Wada, Y., van Beek, L. P., van Kempen, C. M., Reckman, J. W., Vasak, S., and Bierkens, M. F.: Global depletion of groundwater resources. Geophysical Research Letters, 37(20), https://doi.org/10.1029/2010GL044571, 2010.

Wen, F., and Chen, X.: Evaluation of the impact of groundwater irrigation on streamflow in

Nebraska. Journal of Hydrology, 327(3-4), 603-617, https://doi.org/10.1016/j.jhydrol.2005.12.016, 2006.

Wunsch, A., Liesch, T., and Broda, S.: Forecasting groundwater levels using nonlinear autoregressive networks with exogenous input (NARX). Journal of Hydrology. https://doi.org/10.1016/j.jhydrol.2018.01.045, 2018.